OE FORM 6000, 2/69

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

ERIC ACC. NO.	CH ACC. NO.	P.A.	PUBL. DATE	ISSUE
ED 039 388	AA 000 566		May 70	RIEOCT70

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒
ERIC REPRODUCTION RELEASE? YES ☐ NO ☒
LEVEL OF AVAILABILITY I ☒ II ☐ III ☐

AUTHOR
Bregzis, Ritvars

TITLE
Machine Readable Bibliographic Records: Criteria and Creation.

SOURCE CODE	INSTITUTION (SOURCE)
BBB01045	

SP. AG. CODE	SPONSORING AGENCY
RMQ66004	

EDRS PRICE	CONTRACT NO.	GRANT NO.
0.25;1.50	OEC-1-7-070866-4575	

REPORT NO.	BUREAU NO.
Review Series-4	BR-7-0866

AVAILABILITY

JOURNAL CITATION

DESCRIPTIVE NOTE
28p.

DESCRIPTORS
*Automation; *Libraries; *Library Services; Information Services; *Bibliographic Citations; Shared Services; Library Technical Processes; Library Networks; Evaluation; *Objectives; Electronic Data Processing; Cataloging

IDENTIFIERS
*Library Automation; Machine Readable Cataloging; MARC

ABSTRACT The centrality of bibliographic records in library automation, objectives of the bibliographic record file and elemental factors involved in bibliographic record creation are discussed. The practical work of creating bibliographic records involves: (1) data base environment, (2) technical aspects, (3) cost and (4) operational methodology. The application of automated processes to library service functions is dependent on the availability of appropriately structured and functional bibliographic data files. There is a general lack of such files. The known bibliographic record files range widely in their scope of coverage, size, detail of data coverage, functional orientation, and method and cost of production. As a rule they are not mutually compatable. The machine readable bibliographic record services offered by the Library of Congress and the British National Bibliography constitute a trend in distribution of machine readable records of standardized definition and multi-purpose functionality to the library world at consistently increasing rate. Although cooperative creation of large bibliographic record files appears to be a feasible objective for the coming decade, it is not clear to what extent a similar sharing by the small library of the required computing services will become possible for purposes of cooperative utilization of the cooperative bibliographic data files. (NH)

GPO 870-390

ED 039 388

ERIC

Educational Resources Information Center

REVIEW SERIES NUMBER 4

MACHINE READABLE BIBLIOGRAPHIC RECORDS: CRITERIA AND CREATION

CLIS

Clearinghouse for Library and Information Sciences

AA 000 566

LIBRARY SCHOOL • UNIVERSITY OF MINNESOTA

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

REVIEW SERIES NUMBER 4

MACHINE READABLE BIBLIOGRAPHIC RECORDS: CRITERIA AND CREATION

by

Ritvars Bregzis
Associate Librarian
Systems & Technical Services
University of Toronto

This document was prepared pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.

ERIC CLEARINGHOUSE FOR LIBRARY AND
INFORMATION SCIENCES
University of Minnesota
Minneapolis, Minnesota 55455

MAY 1970

The ERIC Clearinghouse for Library and Information Sciences is a joint project of the Educational Resources Information Center of the U.S. Office of Education and the University of Minnesota Library School.

The views expressed by the author in this paper are his own and
do not necessarily represent the point of view of the University of Toronto.

Contents

I.	Preamble	3
II.	Centrality of the Bibliographic Record	4
III.	The Objective of the Bibliographic Record File	6
IV.	The Elemental Factors	7
V.	Functional Criteria of the Machine Readable Record File	11
VI.	Current Practice of Bibliographic Record Generation	16
VII.	Operational Methodology	17
VIII.	The Cost	20
XI.	Conclusion	23
X.	Selected Bibliography	25

I

PREAMBLE

The current dialogue about library automation is characterized by a noticeable undertone which appears to search for plausible, meaningful and rational objectives of this automation. It is almost as if a technology were in search of an application, rather than urgent service needs requiring a more adequate technology.

The various library automation projects over the past seven years have demonstrated a certain amount of technical feasibility of automated procedures in library environment. Along with it they have indicated some exciting potential for these techniques in new areas of library service and at a higher level of effectiveness than is inherent in the customary library procedures even if they are automated Yet little has been accomplished in assessing and harnessing the potential of this new technology in support of the very end objectives of information services and of libraries in particular. One fails to find convincing instances in which automation applications in libraries account for significant new service effectiveness or economical advantages of the library's operational practice. Library automation, with a small number of exceptions, is still tied to the concept of customary library procedures. More imaginative applications, oriented to the end objectives of libraries, still are in the process of experimentation. The end objectives of library service still remain to be defined in terms of specific functionality and economic feasibility

Experience to date has emphasized the attractive potential as well as the demanding aspects of automation, requiring complex technical preconditions and heavy investment of competent personnel and costly machinery. These two factors have prompted consideration of sharing the automation effort and distribution of the resulting benefits. The precise method of this sharing, however, has eluded efforts at definition because of the lack of identification of specific objectives of library service which could be supported by imaginative automation at a high level of effectiveness for the sharing parties

The limits of technical feasibility, the economics, the level of improvement of service, the scope, extent and terms of cooperation, and various ideas concerning library oriented networks, their technological problems and conceptual objectives continue being intensively discussed by various cooperating groups. In spite of this activity, the emergence of practical cooperation in automated bibliographic services and emergence of automated information networks at present are nearly as remote as they were when this dialogue began several years ago. The problem is not merely one of technology; even more basically, the library community must define its specific service objectives.

II

CENTRALITY OF THE BIBLIOGRAPHIC RECORD

Present efforts in library automation are aimed at developing automated procedures, specialized data handling techniques, record storage and access methods, and generation of hard copy records. All these activities depend for their operational success on the availability of sufficiently comprehensive files of bibliographic records of adequate depth and resolution. This does not appear to be always accepted, however, and the development of such files to date appear to be less widespread than development of the various procedures and techniques of automated operation.

Automation of library oriented procedures has been and still is the most popular area of library automation activities. This is a historically logical trend. Library automation began a decade ago with the express purpose of discovering machine aided techniques to expedite the principal business oriented and inventory control procedures. The library literature to date records a great number of mechanized systems aimed at the control of book fund expenditure, acquisition record control, binding record control, serials acquisition control and routing, and book circulation control. With few exceptions, such applications have remained isolated operations as they can contribute at best only in a limited way to the creation of machine readable bibliographic record files, rather than functioning as organic special purpose extensions of a central store of bibliographic records.

Originally, the development of procedure-oriented library systems was based on the total systems concept, popular in the early 1960's. According to this concept it was accepted that the totality of library procedures is functionally interrelated and it was assumed that they should be approached from above as an integral functional whole, where the individual procedures are developed as components of the total system. This objective of library automation was applied to the totality of library procedures as the focal element, disregarding the fact that in library, that is information management, the information resource, rather than the functionality of information, constitutes the focal point of this management.

Over the past years there has been a drifting away from this outlook. The experience of the past decade has demonstrated that in the library environment the bibliographic information-biased complexity defies implementation of the pyramidal procedure-oriented total systems approach. It is being discovered that those library procedures which are presently feasible to be automated have to be built from ground up and that their eventual integration has to be left to the logical confluence of these individual procedures through their common dependence on bibliographic data. The total systems approach as a condition of automation of procedures is giving way to problem-oriented and data-based approaches to automation.

Problem orientation in the library environment is orientation to bibliographic information which in turn is based on bibliographic records and bibliographic data. Hence library functions and procedures involve in some way reference to bibliographic records, to the library "catalogue" as the heart of the total library-housed information transfer mechanism. Consequently, the more intense efforts of library automation are presently leading to the conviction that, with respect to library functions, the level of common reference is found not in procedures themselves but in the store of bibliographic macro-information – the bibliographic record file.

In contrast to this emerging conceptual reorientation there presently exists the reality of general lack of machine readable bibliographic records and paucity of files. Although there are a fair number of projects which involve creation of machine readable bibliographic records, in most instances these records are produced only for a small part of the library's acquisitions, and only on a current basis; that is, beginning with the initiation of the particular project. There exist relatively few machine readable files which cover also the retrospective period of the library's acquisitions, and almost all of these are limited to only the most important data elements of the record. The number of presently existing machine readable bibliographic record files which cover a significant proportion of a library's holdings, and which contain a sufficiently wide scope of data elements suitable for multi-purpose use, is extremely small.

As a result, an imbalance exists between the present attainment in standardization, technological capability, and implemented procedures and techniques on the one hand, and the availability of adequately data-rich bibliographic records on sufficiently large scale in universally usable form on the other.

This is a complex and practically difficult problem for libraries to solve. The bibliographic record is central to the organization and administration of library materials and services. Application of automated techniques to this organization and administration emphasizes anew the central role of the bibliographic record. Library automation essentially is bibliographic file oriented and not procedure oriented. Availability of bibliographic records in machine readable form and provisions for constant updating of the volatile and frequently changing elements of these records are fundamental to successful library automation, even if difficult to attain. What are then the critical aspects pertinent to the creation and maintenance of machine readable records for the variety of purposes libraries appear to want to use them?

III

THE OBJECTIVE OF THE BIBLIOGRAPHIC RECORD FILE

The purposes of the bibliographic record file can range from a simple support of one specific function such as creation of brief machine readable identification cards for books to be controlled by an automated circulation control system to a general multi-purpose file intended to support any of a variety of functions associated with procedural or bibliographic data storage and retrieval operations in either sequential batch processing or direct access on-line mode.

The increasingly widespread accessibility of the third generation computer equipment and techniques is shifting the orientation of data base philosophy toward the more general objectives of multi-purpose systems, on-line access, and inter-institutional compatibility. This shift is being caused by the desire for bibliographic cooperation and a large variety of proposed and planned bibliographic data network systems. It should be noted, however, that highly desirable as such generalized data file objectives are, at present they largely lack precise and proven definitions of scope, depth, logical structure, and level of detail. These definitions with precision are not obtainable without analysis of the logical mechanisms and implementation techniques of such multipurpose inter-related systems. As yet there does not exist sufficient proven knowledge of how such co-operative schemes and networks can be reliably and adequately implemented. Without this knowledge some of the critical requirements for multi-function bibliographic data files cannot be defined.

The purpose of bibliographic records in machine readable form cannot be limited to the customary procedure-oriented general objectives associated with the acquisition, serials control, circulation control and catalogue record handling functions. These functions really are composite, special problem oriented operations geared to a specific environment constrained by the limitations of the existing manual record keeping systems. Application of automation implies a possible change of this environment which may eliminate some of the customary constraints but which may introduce others in turn. Instead of the control of records and the customary associated functions, automation permits and requires control of the component units of records and component elements of the customary functions. This change from purpose oriented records and functions to functional neutrality at the elemental level of the record constitutes the basic potential of the new technology. Instead of information molecules it gives us the opportunity to deal with this matter in terms of atoms, with all the implied benefits and hazards. It provides the opportunity to apply the endless range of transformation formulae created by our ingenuity, but it also carries the risk of potential breakdowns should our formulations fail to account for some of the complex logic in this atomic structure.

Viewed in this light, there are no acquisitions records, circulation records, catalogue records, and serials records *per se.* There are only bibliographic records with their multi-dimensional functions, explicit or implicit. The bibliographic record represents an information item described (descriptive features) and assessed (classificatory, subject-matter features) in a form suitable for communication between the record and the human user. The mechanism of communication is irrelevant as long as the communication is guaranteed. Automation permits, and for effective control requires, that this communication take place always through the basic elements, the bibliographic record. The functional orientation of this communication is represented only as a transient property of the record, as and when needed, without biasing the record towards any specific functionality. Thus, instead of a circulation record, the machine readable file consists of bibliographic records, which are assigned properties pertaining to the circulation aspects related to the items represented by this record as needed. Instead of a specific catalogue record there exists the same bibliographic record with all the aspects pertinent to the correlational look-up functions associated with the "catalogue", inherent in this record.

The basic decisions with respect to creation of neutral, or multi-functional, machine readable bibliographic records therefore are related to the *elemental factors,* as they determine the ultimate serviceability and usefulness of these records in predicted and unpredicted relationships. Such elemental factors are the following: the definitive and structural *formalisms* of the bibliographic record, the definition of the *level* of detail of bibliographic data, the level of structural and *functional control,* the extent of *alphabetic representation,* and the *structural environment* of bibliographic data.

IV

THE ELEMENTAL FACTORS

Bibliographic control utilizes characteristics of bibliographic items expressed as bibliographic data which are formalized in the framework of the bibliographic record. The objectives of this control are varied, and the bibliographic record is required to be responsive to any of the functions relevant in attaining these objectives.

The variety of the objectives of bibliographic control ranges from ability to identify the most specific aspects of a single bibliographic item to the requirement of international compatibility of bibliographic files. The presently prevailing bibliographical conventions have attempted to identify and to define the common denominator level which can be applied to this spectrum of requirements. Attempts to standardize classification schemes, rules for establishing headings, and rules for establishing entries, are the customary methods for this task, based on the prevailing item-per-record, or unit card technology. The level of the common denominator in this system is the bibliographic record, or one of its several dimensions, such as the editorship ascribed to the bibliographic item.

The computer based techniques make it possible to move the common denominator of bibliographic identification towards specificity by about a magnitude. They make bibliographic control possible at the level of the elemental components of the bibliographic record, and in doing so provide additional opportunities for forging new access mechanisms to bibliographic information which are not only more specific but which can also be different in kind from the customarily available access routes. The elemental aspects of the bibliographic record therefore are the principal factors on which automated bibliographic control rests.

Several important aspects of formalization characterize the elemental structure of bibliographic records: the structure of the data elements of the record, the functionality of their elements, and the level of record specificity. *Formalisms* have been elaborated to retain these distinctions in the process of definition of the bibliographic record at the level of its elemental *structural components.* The MARC (Machine Readable Catalogue) record format represents the presently accepted general version of these formulations, and the official acceptance of it by the Library of Congress and the British National Bibliography and by the library community has lent to this formalization the status of a standard.

Systematic identification of bibliographic data elements as constituent parts of the bibliographic record became a central problem during the early attempts to build automated catalogue information control systems. In planning the implementation of these early projects it became clear that the library profession has never been seriously concerned with systematic exposition of the structure of the bibliographic record. In the early 1960's there did not even exist a systematic listing of all data elements that can function as structural components of the bibliographic record. Librarians engaged in the design of the first projects had to do themselves both the analysis and the structural formulation of the bibliographic record (cf. 10, 28, 37). These first implementations and the generally felt need for systematization of bibliographic structures gave rise to the general development of the MARC data format (3). Development of several systems was based on these early MARC definitions, e.g. University of Chicago Library, University of Toronto Library, and Washington State Library. The final report of the Special Project on Data Elements for the Subcommittee on Machine Input Record (SC-2) of the Sectional Committee on Library Work and Documentation (Z-39) of the United States of America Standards Institute was published in 1967. It summarized the principal structural elements of various types of the bibliographic record (17). This first systematic inventory served as the basis of further and more specific documentation of bibliographic data structures and provided the required factual components for the formulation of the MARC II format by the Library of Congress (1), and by the British National Bibliography (4).

The MARC format defines the bibliographic record as consisting of data elements expressed in text form, data elements expressed in symbolic form, data control elements, and elements of technical control of records. While the first two types of elements constitute approximately the scope of bibliographic data found in a customary full catalogue record, the data control elements in the customary record are largely implied. In the machine readable record the data control elements are defined explicitly, while elements of technical control are characteristic to the machine readable bibliographic record alone.

In the process of creation of machine readable bibliographic records the systematic identification of data elements is the principal task, requiring accuracy and consistency. This identification is aiming at technical compatibility of the resulting machine readable records within and beyond the local environment.

Technical compatibility and convertibility of machine readable records facilitated by a standardized record structure, however, does not in itself guarantee compatibility of the record contents. The *data content* of the record is determined by the rules and codes which are applied in the process of formulation of the vital data elements, especially headings, of bibliographic records. It is therefore most important that for purposes of effective exchange of machine readable bibliographic records the technical compatibility of the record format be complemented by the compatibility of data content. Standardized code and its application to the derivation of bibliographic record headings is as vital as standardized application of the norms of machine readable record composition. This is of particular importance in international cooperation and record exchange. The inconsistent application of the Anglo-American Catalogue code to the LC/MARC records through the Library of Congress policy of superimposition therefore is a serious concern for international cooperation(14, p.185).

It is also important to note that the MARC machine readable bibliographic record consists of a) standard data elements which are the major part of the customary bibliographic record, b) alternative elements, that is those which normally are selected one of a kind (e.g. Dewey classification number from several classification numbers), and of c) optional data elements, which may be selected for the records of an individual record file, but which cannot be expected to be universally supplied or accepted. The standard set of data elements can be found complete for a full bibliographic record, or it can be found reduced for a limited bibliographic record. The observance of these definitions makes possible convertibility between individual record files to the level of the least common denominator.

Compatibility depends further on identification of the data control elements and the elements of technical control, consistently identified. This identification customarily is made by use of special coded schemes, and it is essential that, whatever the coding scheme used, the definition of the control data be consistent with the prevailing standard. Thus, the MARC format defines the form of personal name heading in terms of forename, surname, multiple surname, or name of family. Serious incompatibility would arise for the library which would identify the form of personal name in terms of a different categorization.

The MARC record structure provides for specific identification of virtually all bibliographic data elements used in customary bibliographic practice. In addition it also provides for explicit identification of certain customarily implied information. All these data elements are not only recorded and identified as such, but also certain of their specific aspects are defined, usually in coded form. Thus, a specific subject entry is not only recorded and identified as such in easily machine recognizable form, but in addition its specific characteristics are also noted; e.g., that it is a corporate type of name, that it is the direct order corporate name and that it consists of two structural elements, the second of which constitutes a topical form subdivision; as in the case of "Special Libraries Association – Bibliography".

Projection of the general implications of these examples onto the entire bibliographic record indicates readily that the full bibliographic record in machine readable form contains a considerable amount of definitive structural information which increases the volume and the complexity considerably beyond the customary catalogue record. For this reason, creation of full bibliographic records in machine readable form is a complex, time consuming and expensive process, requiring meticulous attention to structurally sensitive detail. Approaches and methods to overcome these costly aspects have been one of the most explored aspects of record generation. The result of this concern is a variety of tried approaches and a variety of levels of coverage of bibliographic data in the record.

Search for the economically as well as functionally acceptable structure and completeness level of the bibliographic record has largely been rationalized by the character of immediate use of the machine readable records. Bibliographic records for the support of inventory control system require relatively few data elements. Production of look-up listings without detailed bibliographical features requires only a limited number of data elements to be included in the record. On the other hand a machine readable bibliographic record file to be used for a variety of needs, such as the production of hard copy catalogue records will require almost the full set of data elements included in the MARC record.

In practice, machine readable record generation ranges across the full spectrum of this structural complexity. While probably one of the most concise bibliographic records produced is that created by the University of Rochester Library for a short title catalogue containing four data elements per record(38), at the other end of the spectrum is the augmented catalogue record produced by Project INTREX(6), containing up to 115 data elements. The MARC record provides for over half of the latter number.

The *functional potential* of data elements is the second important aspect of the elemental structure of the bibliographic record. Like the verb in the sentence structure, the functions associated with the data elements of the bibliographic record define the relationship of these data elements to the environmental purpose of the record. The functionality defined in the designation "author", "compiler", "writer" or "translator" establishes the particular functional relatedness of the bibliographic record containing this name to these degrees of authorship, and between this record and other records which share this name but not necessarily the same function.

This functionality is not unknown in the structural concept of our customary catalogues as manifested in the functional scheme of the entry system. The main entry and the secondary entry, the added entry and the subject entry as two kinds of the secondary entry are the well known aspects of the functionality of bibliographic data which attribute purpose to the existence of bibliographic records. In a machine controlled system a sound scheme of bibliographic data representation and identification is required to account for the definition of this functionality without affecting the identification of the structural elements of the bibliographic record. To take one prominent example, the structural identification of the machine readable record is required to define the "heading" independently from the "entry". Where the "heading" represents the data structure, the "entry" assigns to it the functional property of serving as an access point.

In computer controlled bibliographic record file handling this distinction is fundamental. It provides the powerful capability to select and align records for any desired function independently of the inherent characteristics of the record. It permits to consider as entry any data element which may serve as suitable access point to bibliographic records in a record file.

The structural components of the bibliographic record may be composite structures in their own right. In these instances yet another elemental factor comes into play: *the bibliographic level.* It requires specific identification. For instance, the data element containing "holdings" listed in a bibliographic record may be structured to account for the varying patterns representing the bibliographic and physical aspects of the items covered by the data describing the component parts of the holdings.

In the generation of machine readable bibliographic records, structurally correct identification of the level of bibliographic data elements is of the utmost importance, if systematic access to the records through these data is to be assured. The distinction between the level of series or serial, the level of monograph, and the level of "analytic" is required in order to properly correlate a given data element, such as a personal name relating to the various levels in specific instances. "John Robertson" as editor of a series is associated with the series level. The same "Robertson" has also authored a book which is defined at the bibliographic level of monograph, and he has contributed a paper to a symposium in which his contribution is defined at the "analytic" level. In a bibliographic record file, depending on the search objective in a given instance, it frequently will be convenient if not necessary, to isolate, for instance, the latter two contributions from the first in which our author has likely not contributed as a true author.

The extent of the *alphabetical representation* of the bibliographic record is yet another elemental factor to be accounted for in future-oriented record generation. Along with the increasing sophistication of the computer equipment and its operational logic the ability to extend the repertoire of symbols which can be used in recording of text has grown from year to year.

Up to the early 1960's, computer processing of textual information was limited to upper case Roman letters and a small number of special symbols. Notwithstanding the arguments for and against the adequacy of textual information representation by upper case letters only, the need to preserve the textual characteristics of bibliographic data spurred the development of facilities using the complete character set of the Roman alphabet.

In bibliographic practice the full Roman alphabet and the special and diacritical symbols were first introduced in 1963. The advent of the third generation computers using eight binary bits for character coding made this facility more readily available, so that at present a full alphabet is considered standard for bibliographic data representation. The specifications of the MARC II format provide a series of graded compatible character sets of 64, 128 and 256 characters respectively. While the 128 character set is following the United States ASCII standard, the 256 character set represents an information systems oriented ASCII extension. Presently operated bibliographic record generation projects, with only few exceptions, employ a character set of at least 88 characters, which provides for both the upper and lower case characters, some special symbols and a fair number of diacritical characters used with the Roman alphabet.

For many libraries the problem of alphabetic representation of bibliographic record does not end with the Roman alphabet. Research libraries acquire a large number of publications in non-Roman languages, and current cataloguing practice requires that records representing these publications, for reasons of accuracy, be written in the alphabet used on the publication. The result of this practice is the existence of a large number of bibliographic records produced by research libraries in Cyrillic, Arabic, Hindi and other alphabets or ideographs.

To date all known projects including encoding of such records into machine readable form use transliterated or Romanized equivalents for machine data input. This method involves two problems. One of these is imprecision induced by only partial one-to-one relationship between the values of alphabetical symbols of any two alphabets. The other is the requirement to produce at the output the needed text of the bibliographic record in the customary alphabet of the original language and alphabet. By itself the technical capability to generate a non-Roman alphabet cannot completely recreate Cyrillic or Hindi symbols from the originally transliterated input. This is again due to the various inconsistencies between the phonetic symbol systems of the individual alphabets.

For these reasons vernacular alphabet input is most desirable for representation of bibliographic text in cases where precision and recreation of the original bibliographic text is important. In the research library environment it is important, and recently some effort is being devoted to the development of methods of non-Roman alphabet representation for input of bibliographic records in machine readable form.

Finally, an elemental factor of critical importance for effective functionality of machine readable bibliographic records is the *structural environment* of these records.

In a functional library bibliographic records do not exist individually by themselves. They are the integral units of an information system which is structurally tied together by a number of systematic networks of normalization and control. The most important of these are the networks of names, of topical terms and of classification-shelving symbols.

The current practice of composition of bibliographic records is based on rules determining the choice and form of record headings, of topical terms, and of classification symbols. In the application of these rules the interpretive action of individualistic human competence becomes the critical factor. The result naturally is arbitrary and less than systematic. The three control and normalization networks have been evolved to overcome at least partially this element of individualistic interpretation and its arbitrary results. In addition, these three networks function to provide some redundancy toward the unpredictable choice of and synonymous use of the chosen access data: headings, terms and classification symbols.

In an automated environment these control and normalization networks become even more essential and more demanding in terms of their scope and precision. Bibliographic record files in an automated systems environment are inconceivable without well functioning and precise referral mechanisms in the key areas of control-sensitive bibliographic data. The effort required to ensure the critical level of precision of those mechanisms in automated systems is far more demanding than is the case in customary library record systems. This is one of the reasons why strictly speaking there exists to date not a single functionally integrated automated control mechanism of the kind described.

With a view to the future, however, it is essential that provision be made for these essential control networks of names, topical terms and classification symbol structures. In practice this implies definition and expression of the corresponding data elements in the coded record in such a way that these can be readily related to the pertinent elements in other files and thus establish a function parallel to the "authority" record files in the customary catalogue systems.

V

FUNCTIONAL CRITERIA OF THE MACHINE READABLE RECORD FILE

The foregoing discussion of the *centrality* of bibliographic records in library automation. the *objectives* of the bibliographic record file, and the *elemental* factors involved in bibliographic record creation, provides a conceptual frame of reference. The practical work of creating bibliographic records involves a number of other specific concerns related to the data base *environment, technical aspects, cost,* and operational *methodology.*

Data Base Environment

Bibliographic records exist in a dynamic environment where the elements of these records, their specific controls and relationships with other records are subjected to a constant change. The scope of bibliographic records in a given file may be limited to a single type of records, e.g. monographic publications, or serial publications, or maps. Or, it may be general, including a variety of such types. The file of bibliographic records may be independent and self-contained; or, it may be related to files of similar scope and structure; or, it may be integrated with supporting files which control certain key data elements in all records in the principal file. The complexity created by such file inter-dependence or supporting file structures may be considerable.

This complexity is further underlined by the requirement for continuous expansion and change in the records. As technical capabilities improve and opportunities arise, it is necessary that the creation of bibliographic records take advantage of:

- more effective and suitable input methods and procedures
- newly available technical facilities, such as expanded character sets or additional alphabets
- expanded scope of the record creation effort, e.g. adding records for specialized materials such as maps, or, extending the activities to retrospective record creation
- expanded coverage of the detail in the records, for instance, upgrading the records from skeletal coverage to a fuller form, such as the "MARC record".

Technical Aspects

In addition to the elemental factors of data and the conditions of the data base environment there are numerous technical aspects which are important in the process of creating machine readable bibliographic records.

Unless the bibliographic record is structurally simple, a specially prepared *source record* is required for machine readable record creation. Depending on the specific requirements, the readily available bibliographic data and procedural convenience, any of a number of methods may be selected, ranging from ready "editing" a customary catalogue record, to photocopying combined with elaborated editing of a source record, to filled-out data sheets specially designed to accommodate the required explicit data along with explicative annotations and functional coding.

Encoding, likewise, may employ any of a number of available techniques: Hollerith cards keypunched on a keypunch machine, paper tape typewriter, magnetic tape typewriter, or typing into a directly computer controlled medium such as disc or other buffer storage. There are numerous details involved in each of these techniques. Some of these may affect the ways of utilizing machine systems, while others may be important in relation to the effectiveness of input and to human convenience.

Another important consideration, frequently taken for granted if not forgotten, is that bibliographic records are highly dynamic and that *change* in some of their data elements is the rule rather than the exception. This requires a technically efficient and rigidly controlled update mechanism built into the entire process of bibliographic record generation, beginning with update data verification, then following through keying, addressing the pertinent records, verification of the update validity, through to the incorporation of the update data in the pertinent record. And all of this under rigid control, in order to safeguard the *integrity* of the machine readable data file.

This latter is a particularly difficult and critical task. While the machine readable record file generally grows over a period of time, some aspects of the record conversion operation unavoidably change, or are subjected to planned alterations. Staff changes, adjustments in data definitions, changes in supporting computer programmes and changes in required formats for verification and control are taking place. Even under the most rigorous system of documentation of such changes, unforeseen inconsistencies in the data structures on the file develop long before they are noticed, to say nothing of slips in human consistency and machine performance. The result of these slippages is, over a period of time, emergence of trends of practice which are not covered, or are covered inadequately by the existing documentation, however meticulously kept. Existence of these unexpected characteristics in the file usually is first recognized when the entire file is tested for some specific new use. It is then a labour of ingenuity and patience to track down, diagnose and to revert these illegitimate factors to their legitimate and controlled form.

The file integrity problem is further complicated by lack of complete parallelism in the definition of data elements in *records of different origin.* As the file of machine readable records grows, it accumulates records originated by a variety of sources. To begin with, in-house produced machine readable record files tend to grow more complex formally and richer in detail as experience of the conversion process is accumulated and as more effective ways of converting are devised, resulting in record upgrading from time to time. It is not always practical to implement this margin of improvement immediately on the already converted portions of the file. Further, as more sources of bibliographic record distribution become available, machine readable records obtained from external sources by exchange or purchase result in increasingly varied collection of records. Normally it is not feasible to bring all the newly acquired records to the accepted standard immediately at the time of their acquisition. Subsequent expansion of the data is usually a complex task, particularly if the expansion pertains to different levels of detail over a large file.

In practice it is necessary to provide for continuously controlled and reliable upgrading of machine readable bibliographic records. The level of definition detail of bibliographic records has been on the rise over the past five years. Beginning with customary catalogue data in 1963, the library profession has elaborated the MARC format standard of bibliographic data definition, which includes a considerable amount of bibliographic data over and above the customary catalogue record, and which has become the *de facto* standard of machine readable bibliographic data definition. The practice of machine readable bibliographic record generation reflects in proportion this growth in data level, and it is not uncommon to find that a record generation project is upgraded from time to time in order to assure increased and more versatile capabilities of the resulting data base. To exercise unerring control of the gradually growing record file under this additional dimension of change is another vital and technically critical aspect of the record generation process.

One dimension of such record upgrading merits particular attention: the desired increased sophistication in use of *alphabetical symbols.* Use of the full set of Roman alphabetic symbols has become a generally accepted necessity for accurate bibliographic definition, compared with the early 1960's when lower case letters were accepted as unavailable for text processing. The present technical feasibility permits use of even larger character sets, and the recently defined MARC character set lists 142 alphabetic, numeric and special symbols for use in the Roman alphabet. This character set, however, does not represent a practical technical limit at this time. Presently methods are being developed for encoding also non-Roman alphabets, e.g. Cyrillic, Arabic, Hindi. The recent development of computer output microfilm (COM) techniques permits printing of a large variety of type font mixes, permitting computer generated printout of a catalogue displaying records in Roman and other alphabets. These developments can be expected soon to be reflected in record generation practice, which would require some further complexity of the technical aspects of record encoding.

Encoding of the individual characters and character strings along with the identifiers which should provide these capabilities thus becomes more complex and requires more binary bits for discrete identification. The *8-bit character* which now is the standard with most of the third generation computing machinery can serve these requirements reasonably, although not often elegantly. A 6-bit character machine, however, is required to be used ingeniously and therefore less flexibly, indeed, to accommodate the extensive character manipulations of these large character sets.

In the process of machine readable bibliographic record generation, the computer, the *operating*

system, the application programmes and a variety of related matters can be not only difficult but also frustrating. Generation of machine readable bibliographic records implies very large files, extremely variable and highly complex structured record format, and large character sets. These characteristics further imply large scale storage devices, operating systems that can support very complex sorting facilities of multi-tier structured records, extensive character and character string manipulating capabilities, and refined printout facilities. These are not requirements readily available from a typical computer installation. To obtain these it is necessary to invest large resources, considerable effort and high level talent. The result of such effort frequently is modification or extension of the operating system and utility programmes. Although devised to accomplish the specific required operations, at the same time they tend to create clashes with subsequent editions of the manufacturer's operating system and utility programmes, with the resultant difficulties for the management of the computer installation.

There are further concerns and problems related to the *application programmes.* In most cases they have to be written for the specific bibliographic record generation system for a specific computer configuration. Frequently, no sooner have these programmes been tested and put in successful operation, than changes in the computer hardware configuration, operating system or some other vital area take place. This necessitates adjustments or even modifications in the various applications programmes. Such changes are prone to leave behind them unexpected traces in the record file.

There are ways of guarding against many unexpected twists in the record files created and maintained by the complex computer controlled processes. The definition of these, however, involves a most detailed understanding of both the bibliographic logic and the logic of the software of the computer system. This is difficult to attain and reliable *data validation procedures* in present practice are the exception. (For some advanced methods of error detection and quality control cf. Dolby *et al.*, ref. 22, p. 71-83.)

Another aspect of record safeguarding involves a variety of file *protection* features, ranging from the trusted method of redundancy to various more or less complex logical procedures involving severe control of update access to records, and monitoring and auditing of all changes made. A reasonable redundancy is probably one of the most reliable kinds of assurance that can safeguard against data loss arising from equipment malfunction, software failures, application programme errors, procedural flaws, human errors, and a great many unexplained matters that go wrong. A rigorous procedure of keeping a backup copy of every file at least until the next processing phase is obtained, is a method which frequently repays its cost.

In building and maintaining the bibliographic record file, a number of *data manipulatory provisions* are essential. Systematic, explicit and non-ambiguous *definition and identification* of data elements in bibliographic records is essential to permit controlled logical manipulation of bibliographic data. On the precision of this identification depend all the bibliographic control functions which constitute the principal reason for invoking the tool of automation in this effort. Control of the integrity of bibliographic records, precision of their transformation, augmentation and printout of any desired pattern of sequencing of bibliographic record files, and the relevancy of on-line accessed bibliographic records, all depend on the precision which has been devoted to the data definition and identification in the machine readable bibliographic record files.

One of the most important procedures which require certain control data elements built into the bibliographic record is machine sequencing of bibliographic record files. *Machine filing* algorithms depend on identification of all bibliographic data elements which can be made functional in obtaining the desired filing sequence, as well as on those which constitute exceptions.

For technical reasons frequently it is practical to maintain the machine readable bibliographic record file in several forms so that certain unlike groups of operational functions can be readily accommodated. Thus, for the purposes of a circulation control system or of printout of brief accessions listings only a limited number of record elements are required and in certain situations it may be preferable to maintain appropriate specialized files with only the required data elements. It is essential that the machine readable bibliographic records contain definitions which permit such and similar abstraction of data for the creation and independent maintenance of such *subsidiary files.*

The methods of computer use for bibliographic data handling likewise are vital and appropriate provisions are necessary in the definition and identification of bibliographic records. *Batch processing* procedures which depend on predefinition and cyclic performance of all processes of data manipulation,

while generally less demanding in definition of data control elements, are also less tolerant of inconsistencies in the data structures. An incompletely defined record will frequently be disregarded in a batch processed operation.

Bibliographic record use in *interactive* on-line processing procedures demand less functional predefinition of the processes and therefore are more tolerant of some formal imprecision, at the expense of more precise elemental identification. Other kinds of imprecision usually do not cause the record to be disregarded in processing, but instead can refer this imprecision to the on-line user who then is given the opportunity to instruct the machine system to apply an alternative control aspect for similar acceptable results. Thus, search of a specific bibliographic record without sufficiently precisely known title of the publication submitted to a standard batch-processed search programme would normally be expected to be unsuccessful. In an interactive search process the search may be continued by using alternative data elements, e.g. language, decade of publication, or others, to reduce the set of likely candidates; this would likely result in a small number of records, among which, upon viewing these, one could recognize the record being sought. Data elements of finer cut can therefore be of important use in various cases of marginal definition.

In the actual use of machine readable records, characteristics of computer file devices, techniques used for record *storage*, indexing and *access*, and processing conditions relating to the use of record files are vitally interrelated with the type of data, their definition, volume and encoding patterns. Record use can be facilitated greatly by provisions contained in bibliographic records which permit such functions as precise and logically graded indexing, or efficient file regeneration.

Searching of pertinent bibliographic records is fundamental to most bibliographic data operations. This is also the most difficult operation for the computer to perform; difficult because of the complexity and incomplete knowledge of the underlying logical structures. Effective searching of large text-like files is still an unresolved intellectual problem and new approaches and methods are being tried. All of these have to depend to a large extent on algorithmic procedures. Precision of these depends heavily on the finesse of definition and precision of the data base. In the last analysis it is the potential built in the bibliographic records which governs the effectiveness of searching procedures.

Effective use of bibliographic records in machine based operations depends on *statistical* information for control and improvement of these operations. For this reason as well as for the monitoring of use patterns, it is important that bibliographic records be defined in a way which permits statistical analysis involving any of the important data elements.

Provisions in bibliographic records for effective, versatile and high quality *printout* are obviously most important. It was this provision which caused the greatest concern in the early days of computer applications to library operations, when the principal role of the second generation computer was relatively simple sequencing and output printout operations. Present day technology has advanced somewhat since that time. Full Roman alphabet with upper and lower case letters can be readily printed with most third generation computers. Computer output microfilm devices (COM) capable of large font sizes for any alphabet are becoming available for specialized large scale applications. It can be expected that in the coming years the principal limitation of computer output of text will be restrictions accepted in the original encoding of bibliographic records.

Bibliographic records encoded in upper case letters cannot be expected to be readily transformed into full Roman alphabet text without substantial human assisted editing. Transformation of transliterated Arabic into vernacular Arabic alphabet cannot be expected to be accomplished without human editorial assistance. Graphic quality printout depends on certain typographic function identifications in the text. Not all of these can be generated automatically from the bibliographic text without editorial assistance if appropriate identification of data elements has not been built into the record originally.

Not all data in the bibliographic record require to be printed. It is essential that discrete units of data can be identified for printing, so that the elements to be printed can be recognized by the automatic process, that the required lengths of the units are submitted to the printing process, that appropriate spacing is inserted, that line length can be properly measured, and that the proper type font is selected, so that the high quality printing can be machine controlled.

This overview of technical considerations which affect the use of machine readable bibliographical

records indicates that the usage return of the invested effort depends on the detail of unambiguous definition and identification of the component units of the bibliographic record and the various special characteristics of these units related to the principal functions of file creation, maintenance, access, sequencing and printout generation. To satisfy all these requirements to an adequate level has been the objective of the standardized definitions of the MARC format. To what extent these provisions have been and are being implemented in individual instances of machine readable bibliographic record creation, depends on the objectives, specific purposes, availability of resources and many other factors. The long term value of the converted records however cannot exceed the level of detail and effort that is devoted to the creation of the records.

VI

CURRENT PRACTICE OF BIBLIOGRAPHIC RECORD GENERATION

The considerations relating to bibliographic record generation in machine readable form discussed in the foregoing chapter are demanding in their implementation. It is perhaps largely for this reason that to date there exist only a small number of large and detailed bibliographic record files. The current practice shows a wide variety of scope of definition and detail encoded in machine readable bibliographic records for a varying array of special purpose operations or varying levels of generalized use.

This variety may be viewed in two ways. On one hand it represents files of bibliographic records ranging from abstracts to indexing records of books and serial publications. The difference between these may be viewed as one of specificity rather than one of kind. This discussion, however, does not deal specifically with machine readable abstract and indexing records.

On the other hand the variety of machine readable bibliographic records may be viewed as a spectrum of detail and definition practiced in machine readable bibliographic record files. This discussion attempts to point out the principal variants in this spectrum, which ranges from sketchy records containing only several data elements and virtually no definitive information, to the level of full MARC record, and beyond.

The operational practice of machine readable bibliographic record creation indicates two principal situations in which these records are generated: current record generation, and retrospective record conversion.

Current record generation takes place in a variety of operational and procedural situations. It may be the encoding of a brief record as part of book order procedure, or keyboarding the essential data elements from currently prepared catalogue records for purposes such as circulation control. Alternatively the machine readable record may be obtained through catalogue card typing, or it may originate as the result of full scale bibliographic data preparation on special data sheets used for catalogue record typing as well as for keyboarding in machine readable form. Or the compiled bibliographic data may be edited for machine input and then used for generation of hard copy records, and for addition to the machine readable file.

Any one of these and other similar gradations of operational practice may be found applied to a limited selection of records or to records of all currently acquired and catalogued materials. Encoding of bibliographic records for specialized materials and materials in non-Roman languages is not yet widespread.

Creation of machine readable bibliographic records on a current basis is being practiced by a number of documentation centres, libraries and several cooperative projects. A strict definition of this process is somewhat hindered by two factors. First, machine readable records are not infrequently generated from initially available partial information without resorting to the appropriate procedures of bibliographic description. Records generated during the process of acquisition of library materials fall into this category, and the result may not be bibliographic records strictly defined. Second, bibliographic data are occasionally encoded for a specific purpose, which determines the scope of data elements included in such records. Machine readable records generated solely for identification of books in a circulation control system, or records encoded specifically for production of a book form catalogue with abbreviated listings are examples of this latter orientation. In both cases the resulting records may not be considered bibliographic records capable of satisfying the customary functions of bibliographic control of library materials.

Although most of the existing bibliographic record files are essentially files of records with such limited bibliographic capability, there are at least several projects which create machine readable bibliographic records according to the MARC standard. The MARC record services of the Library of Congress and of the British National Bibliography produce such machine readable records in the MARC exchange format.

VII

OPERATIONAL METHODOLOGY

The Logical Processes

Machine readable bibliographic record generation entails a number of human activities and mechanized processes. Individually they may be simple and undemanding; in their inter-relationship they are critical and their interrelated precision affects vitally the quality and functional potential of the generated records.

Creation of machine readable bibliographic records involves the following principal logical processes:

Obtaining of the *source document* from which the bibliographic data are transferred to machine readable form. This process may involve copying, transcription, annotation, editing and like functions. In its complexity it may range from simple annotation of a title page to elaborate composition, verification, and display of bibliographic data on a specially prepared data sheet.

Editing of the bibliographic data available in the source document. This process secures the appropriate explication of bibliographic data according to the formalisms and definitions of the machine readable record format. This is a critical process as the precision of this work is reflected in the machine readable record. Usually an expert editor, with complete familiarity with bibliographic record conventions, formalisms and logic as well as appreciation of the machine oriented systems implications, is required in order to perform this function acceptably.

Edit revision is the process intended to exercise quality control over the editing of bibliographic records for machine input. In practice this process may range from an outwardly not even visible operation carried out by the editor, to elaborate procedures designed to achieve uniformity of definitions and explications of implicit information. In larger scale operations, particularly where machine readable bibliographic records are generated by several functional units, e.g. current record generation and retrospective record conversion, a consistent operation of this process is of vital importance to securing compatible bibliographic records.

Keypunching is the typically familiar process of machine readable record generation. Although frequently it is understood to be the principal process in such work, in many instances this is not so. Effort and costs attributed to keyboarding usually is the smaller part of the total effort of record production. The emphasis which is placed on the specific method of keyboarding, or the method of data entry in general, therefore does not weigh as heavily in the total effort of record production as is widely held.

Apart from optical character reading (OCR) methods which at present are still generally unsatisfactory for meaningfully significant data conversion, data entry is obtained by one of a variety of methods of keying: keypunching of Hollerith cards, keying on paper tape typewriter, keying on magnetic tape, or keyboarding data into some form of intermediate storage controlled by computer. This latter method, which is available with on-line access to computing service, has the advantage of directly connecting the keying process with the proofing process, bypassing some of the cumbersome pre-proofing acitvities. The major handicap in the keyboarding of bibliographic data experienced with most of the prevailing data keying devices is their lack of parallel hard copy writing facility. Bibliographic data edited for machine input constitute highly complex text, and touch typing without simultaneous sight verification by the typist usually does not produce adequate results, thus placing unduly heavy and unnecessary load on the proofing process. In larger scale operations the simultaneous parallel printout facility can be crucial.

The *pre-proofing* procedures involve a variety of activities ranging from maintaining control over source documents pending the final acceptance of the fully proofed machine readable bibliographic record, to organizing the listings of machine produced proof copy, to maintenance of appropriately phased queuing order in the records requiring additional adjustments or updating. In an operation of any significant size this is a complex task requiring meticulous observation of detailed patterns of operational sequence of source documents, queues, human work and computer processing.

Proofing is a largely repetitive process imposed on a bibliographic record undergoing transformation to machine readable form. The extent and complexity of this process may range from literal proofing of the

machine produced copy against the source document, to detection and analysing of results which have arisen from interaction of input errors with certain machine processing functions. This is the stage at which the most inconceivable errors in the supposedly fully debugged supporting computer programmes are detected in an indirect but most effective way. Understanding of the supporting programme logic is therefore of significant assistance to the proofer.

Post-proofing procedures cover the various record maintenance, collating and quality control procedures required for the final phase of the mechanized record transformation. It is this phase which determines whether the record, having been proofed and again inspected, will be declared a valid machine readable record, or will be cycled back for further adjustment and another proofing. Since normally most records circle this loop more than once, it is important that appropriate controls and quality checks be built into the operation of this process.

The final but perhaps the most complex and hazardous is the process of controlled posting of the created machine readable bibliographic record onto the file of these records--the process of bibliographic *record file maintenance.*

As integration of a catalogue record into the customary library catalogue is a complex process not to be confused with the simple act of filing this record in the catalogue file, so is the integration of a machine readable bibliographic record into a machine readable bibliographic file. And more. Not only must the existing and potential bibliographic complexities related to the constantly changing environment of entries and other formalisms be accounted for, but likewise many aspects of logical and technical control over the machine oriented file in its totality require meticulous procedures to guard against inconsistencies with potentially catastrophic consequences.

The bibliographic record file is extremely dynamic in its nature. The rate of required adjustments in bibliographic record is high. Any record may require updating of some of its data fields at some time. Many require frequent or periodic updating. The updating information may be of varying levels of completeness at the time of updating. A certain level of up-to-dateness of one record may be obtained with one update while in another record the same level is taking several updates.

Bibliographic records in a large file are the result of activity over a period of time. During this period, unavoidably, some methods of record generation are changing, some procedures are changing, and the extent of data in the record may be increasing or decreasing at a certain point in time. For instance, a record which has been created at a time when data were recorded at level A, is subsequently updated at the time when data are recorded at a higher level of complexity, B. Within this update there actually are two logical updates, the change in up-to-dateness related to the two sections of the file, and the implied difference between levels A and B. And this applies potentially to every record in the file. Such inconspicuous matters frequently slip unnoticed by the staff and unaccounted for by the computer programme. In spite of meticulously attempted definitions and documentation of the programme logic and procedure, any most innocent change introduced in the programme is likely to produce unexpected side effects at some time. It almost always does.

The control of the integrity of the machine readable bibliographic record file is a most complex, demanding and costly process. Yet it is a fundamental and inescapable precondition for any effective operational use of the file. Without this control the machine readable bibliographic records remain only elements without cohesion that lends power to the files as a resource.

Administrative Factors

From the operational point of view, machine readable bibliographic record generation involves three distinct key areas of administrative concern. These are: the operational organization, the creation and maintenance of the required computer programmes, and the computer operations supporting the human effort

The *operational organization* of machine readable record generation varies in some of its aspects depending on whether it is aimed at generation of machine readable form of currently obtained or originated bibliographic records, or at conversion of accumulated retrospective files of bibliographic records to machine readable form. The variations are related mainly to the need to correlate current record generation to the current activities of cataloguing, record creation and updating. The sensitivity of this

correlation is heightened by the fact that the logical processes of machine record generation do not naturally correspond to the sequence of processes involved in catalogue record origination. Retrospective conversion, on the other hand, is not characterized by a high level of operational dynamism and requirement for such procedural synchronization.

Under both conditions, current and retrospective, the characteristic of record file dynamism prevails. The operational organization of the machine record creation effort has to cope with the problem of omni-directional and dynamic correlation of the various phases of the effort. Aspects of particular attention include:

- the lead time between the planning and the closing of decisions related to adjustment or revision of processes and procedures; balancing of staff effort; accounting for the required training and conditioning; and multi-stream work flow co-ordination.
- correlation between levels of data complexity and precision in the total input process and in the varying file segments, between the various phases of the project and between different projects.
- intermediate use of the unevenly growing machine readable bibliographic record file for currently required service purposes. This is a heavy requirement, since it imposes some of the demanding requirements of a functional data base upon the file which is in the process of construction and augmentation.

The creation of machine readable bibliographic records is heavily dependent also on *computer programmes* which support the work of the human effort. Input data validation, analysis of data structures, posting of new records, record updating and change, file maintenance, and programmes for printout of proofing, editing and update results, and statistical review, are the principal functions which depend on reliably and accurately operating computer processes. These require considerable effort to plan, design and to code as well as to maintain and adjust so that they reflect the current logical processes of the bibliographic record generation effort.

Bibliographic record generation is an effort which must be sustained regularly. This predicates not only continuous procedures on the part of the staff, but also continuous scheduled operation on the *computer processes.* The continuous or periodic validation of keyboard data, the timely production of proof copy, the controlled addition of new records and changes to the existing accumulation of records are vital operations. On the timely service of these depends the continuous work flow of the proofing, record handling and file maintenance functions of the staff. In practice the guarantee of such uninterrupted computer service is one of the most critical factors in the task of machine readable record generation; it usually is easier, although with increased risk to quality control, to adjust the human effort to meet the situation in hand than to adjust the computer operations when the computer facility is handicapped or overburdened.

VIII

THE COST

As in other aspects of computer applications to library oriented processes, there is only scant and generally indicative cost information available regarding machine readable bibliographic record generation. Whatever cost studies exist, they cannot be easily related to a single pattern and uniform behaviour of the cost factors covered. The reasons for this situation centre around several groups of problems: the variety of bibliographic records; the variety of their environment and structure; the variety of combinations of techniques used for the record generation; and the absence of a standard way of defining and of accounting for the various phases of the record generation process.

In reviewing the several dozen reports of the more widely known projects which include or concentrate on bibliographic record generation in machine readable form, one notices the lack of uniformity with respect to the record content, ways of data identification, methods of encoding, and procedures of operation of these projects. Accordingly, whatever cost data are given in these reports, they usually do not cover sufficiently comparable cost factors, nor are they usable without considerable normalization and prorating for the purposes of meaningful relative cost assessment. Two recent studies have attempted to research the available cost information in some detail. The study by Dolby and others analyzes in detail five cases reported in 1965 and 1966 and shows the average per title cost to range from $.37 to $1.31 (22, p. 40-45). The composite average cost for records of approximately 250 characters in length is calculated at $.48 per title, while for the 425 character long records it is $.90 per title. The authors conclude that "the number of variables in even as well defined an operation as data conversion is so great as to prohibit the construction of a set of equations that will apply across all libraries" (22, p. 43).

Another recent and thorough analysis of this complicated cost situation can be found in the *Preliminary Report on the Review and Development of Standard Cost Data for Selected Library Technical Processing Functions*, prepared by the Information General Corporation (9). Under the rather general title this report deals solely with the conversion of catalogue records to machine readable form. It analyses cost data given in 35 reports covering 22 projects, published in the period from 1963 to 1969, with 1968 imprint date predominating.

The conversion, that is the generation costs of machine readable bibliographic records, according to this analysis ranges from $0.064 to $3.854 per record. At one extreme the coverage is limited to the call number and information identifying the physical item, at the other the augmented bibliographic record includes information far beyond what customarily is called "full cataloguing". The cost of what can be termed normal complexity bibliographic records ranges from about $.40 to about $2.00 per title depending on the method of conversion, extent of explication of implicit information and manner of operation.

The experience of machine readable record generations reported by the following institutions is of particular interest for the individual assessment of the variety of factors involved: Los Angeles County Public Library(34), Johns Hopkins University Library(25), LC/MARC Pilot Project(39 and 2, p. 67-76), Michigan State University Library(12), Ontario New Universities Library Project(10), Project INTREX(6), Purdue University Library(23), University of California at Santa Cruz(8), and Yale Medical Library(28).

The resulting cost of machine readable bibliographic record generation is highly dependent on the variety and extent of the principal factors involved in the creation of such records. Contrary to the widespread assumption that keyboarding constitutes the major part of machine readable generation or conversion, there are a number of more demanding and costly factors which influence heavily the true cost of machine readable records.

Assuming a given copy of bibliographic records, such as the main entry card, shelf list unit card, or LC card, there are two principal groups of cost factors, leaving aside special cost overhead factors such as housing, or amortizing of machinery. The first of these major groups comprises the cost of the operation staff, covering the security of the source document of the record, record editing, revision of keyboarding, maintenance of pre-proofing records, proofreading, maintenance of proofing records, converted record file maintenance, and direct supervision. The other group of cost factors includes the producing and

maintaining of the required computer programmes, the systems operation cost, computer time, and minor equipment, services and supplies.

Among the cost factors of the first group the principal ones are the securing of the source document (which frequently involves copying in some form), record editing for keyboarding, keyboarding, proof-reading, and maintenance of the machine readable record file. The staff time cost for this latter function becomes particularly demanding as the record file grows larger. The extent of editing and proof-reading is directly related to the complexity of the machine readable record, and each of these may exceed the cost of keyboarding.

The cost factors related to the computer systems operations can be particularly varied, depending on the character of the required computer programmes, the extent and frequency of their subsequent modification, the conditions and rate of charges for the computer service available, the type and cost of data input equipment, and any special services, e.g. commercial on-line data entry service.

Customarily, in the published reports of machine readable bibliographic record generation only some of these specific cost factors are identified. Face value comparisons of cost based on the hitherto published reports can therefore be only indicative and approximate. In the total conversion and record accumulation process, quantitative aspects begin to play a noticeable role beyond the one quarter million record file. At present there are only a small number of machine readable bibliographic record accumulations that exceed this volume. No systematic studies of the cost aspects of these accumulations are known to be available to date in the open literature.

The existing reports which include cost aspects concentrate on either limited volume projects or some special aspects of the machine record generation process. It is of interest to note that several projects report data inputting cost around $.50 per record. The machine readable record input for the Stanford Undergraduate Library book catalogue is reported to be $.40 per title(26, p. 27), The mass conversion of 110,000 titles by the University of California Library at Santa Cruz is reported to have cost $.60 per title(8, p. 117) while inputting of 50,000 titles using IBM DATATEXT at the State University of New York at Buffalo cost $.55 per title(5, p. 224). It should be noted, however, that the reported cost refers to inputting of bibliographic records, that is keyboarding, correction and associated direct input equipment costs. Viewed against the preceding discussion of the larger set of supporting processes, this cost is only part of the total expenditure required for creation of machine readable records.

As Bourne and Kassan have indicated, the reported cost analyses are difficult to compare also because they seldom provide base unit costs, such as hourly salary and machine time rates(9, p. 5). Development costs and costs of materials and supplies are not available on a basis consistent throughout all these reports.

Other more subtle factors complicate the problem of cost assessment still further: the average number of characters per record, language of records, the character set used for record encoding, number of data elements in the bibliographic record. The background, competence and experience of the work force, and the specific procedures followed - all affect the cost of the machine readable record generation process.

The inconsistency in cost data reporting and the resulting difficulty of reviewing costs of machine readable record generation is amply evidenced by the foregoing reports and the published accounts of bibliographic record conversion. To a large extent this inconsistency is caused by recording costs according to the procedural steps in a given situation, rather than according to the principal logical functions. Combination of logical functions into a single procedural step often serves to bury specific functional factors in the total cost, thus losing their analytical value for the assessment of the constituent components.

In addition to the salary cost of the procedural operations and the attendant overhead costs, bibliographic record generation and file maintenance imposes cost of computer support programme development and maintenance paid as programmers' salaries, cost of materials and supplies, and cost of computer time. The assessment of these costs is usually rather complex, due to the variety of ways in which these services are secured. However, these cost factors are far from negligible and can significantly affect the total cost average.

In cost assessment it is desirable to distinguish between the three levels at which record generation effort is supported: creation of machine readable bibliographic *records* as un-interrelated units of data, maintenance of the bibliographic *record file* where these records are controlled as units of a functionally

integrated file, and maintenance of the *data base* where the file of records is maintained and controlled within a functionally oriented active use environment. The relative apportioning of the cost of programming, computer operating staff salaries, computer time, and supplies to the creation of records to the maintenance of the file, and to the maintenance of the data base can be of significant importance in cost analysis in large ongoing systems which not only create but also actively use machine readable bibliographic records.

IX

CONCLUSION

Application of automated processes to library service functions in the last analysis is dependent on availability of appropriately structured and functional bibliographic data files. At present there is a general lack of such files compared with the existing procedure oriented automation projects.

The known bibliographic record files represent a wide variety of situations. They range widely in their scope of coverage, their size, the detail of data coverage, functional orientation, and method and cost of production. They all suffer from lack of common definitive and identificatory basis and therefore as a rule are not mutually compatible.

Amidst this diversity a trend toward standardization and multi-purpose functionality has begun to emerge. The definition of the MARC bibliographic record format is on the way to becoming a virtual standard. The machine readable bibliographic record services presently offered by the Library of Congress and the British National Bibliography constitute a trend in distribution of machine readable records of standardized definition and multi-purpose functionality to the library world at consistently increasing rate. There is also some prospect for larger scale availability of retrospective bibliographic records through the Library of Congress RECON project, and possibly similar efforts in Europe.

A number of larger libraries have been engaged in ongoing record conversion for several years and some of these have accumulated record files nearing the half-million mark. In most of these files records are defined as a subset of the MARC standard. Upward adjustment and augmentation is therefore possible. There are a fair number of other libraries which have accumulated smaller number of records in more specialized areas. These too have sufficiently high level of common data coverage to be usable in augmentation toward the standard format and level of definition.

Production of full bibliographic records in machine readable form is costly. However, it appears not to exceed the cost range for cataloguing with Library of Congress catalogue card copy, which in turn is less than half of the cost of original cataloguing. Although precise and generally applicable costs of machine readable record production presently are not available, the approximate range of these costs indicate a definite feasibility of economical production of machine readable records on a cooperative basis.

The ultimate question of the general usefulness of bibliographic records, however, is more difficult to answer. There is no indication that machine readable bibliographic records are directly more economical than their manual counterparts in terms of the customary effectiveness of the latter. On the whole existing machine readable records are capable of supporting enhanced service effectiveness. This support, however, can be obtained only by using costly computer power. Although cooperative creation of very large bibliographic record files appears to be a feasible objective for the coming decade, at this time it is not clear to what extent a similar sharing by the small library of the required and still relatively complex and costly computing services will become possible for purposes of cooperative utilization of the cooperative bibliographic data files. And yet, the potential of utilization of machine readable bibliographic records appears to be too attractive to be left untapped. The need to provide more effective information services and the profit inherent in meeting this need appear to be more than sufficient stimuli for application of human ingenuity to this challenging task in the most creative way. The information community cannot afford not to pursue this task.

X

SELECTED BIBLIOGRAPHY

(1) AVRAM, Henriette D., John F. Knapp and Lucia J. Rather. *The MARC II Format: a Communications Format For Bibliographic Data.* Washington, D.C., Information Systems Office, Library of Congress, 1968.

(2) AVRAM, Henriette D. *The MARC Pilot Project. Final Report on a Project Sponsored by the Council on Library Resources, Inc.*, Washington, D.C., Library of Congress, 1968.

(3) AVRAM. Henriette D., Ruth S. Freitag, and Kay D. Guiles. *A Proposed Format for a Standardized Machine Readable Catalog Record.* A preliminary draft ISS Planning Memorandum number 3. Washington, D.C., Office of the Information Systems Specialist, Library of Congress, June 1965.

(4) *BNB MARC Documentation Service Publications,* Nos. 1 and 2. London, Council of the British National Bibliography, Ltd., 1968-69.

(5) BALFOUR, Frederick M. "Conversion of Bibliographic Information to Machine Readable Form Using On-Line Computer Terminals." *Journal of Library Automation,* v. 1, no. 4, December 1968, p. 217-226.

(6) BENENFELD, Alan R. *Generation and Encoding of the Project INTREX Augmented Catalog Data Base.* Cambridge, Mass., Catalog Data Input Group, Electronic Systems Laboratory, Dept. of Electrical Engineering, M.I.T., August, 1968. AD 677418

(7) BENENFELD, Alan R. "Generation and Encoding of the Project INTREX Augmented Catalog Data Base" (In Carroll, Dewey E., ed., *Proceedings of the 1968 Clinic on Library Applications of Data Processing.* Urbana, Ill., University of Illinois, Graduate School of Library Science, 1969, p. 155-198).

(8) BLACK, Donald V. "Creation of Computer Input in an Expanded Character Set." *Journal of Library Automation,* v 1, no. 2, June 1968, p. 110-120.

(9) BOURNE, Charles P. and Madeleine S. Kasson. *Preliminary Report on the Review and Development of Standard Cost Data for Selected Library Technical Processing Functions.* Palo Alto, Calif., Information General Corporation, 1969. Report No. IGC-PA-69-41.

(10) BREGZIS, Ritvars. "The ONULP Bibliographic Control System: an evaluation." (In Jenkins, Frances B., ed., *Proceedings of the 1965 Clinic on Library Applications of Data Processing.* Urbana, Ill., University of Illinois, Graduate School of Library Science, 1966. p. 112-140).

(11) BROWN, Peter "The Bodleian Catalogue on Machine Readable Records". *Program, News of Computers in Libraries,* v. 3, no. 2, July 1969, p. 66-69.

(12) CHAPIN, Richard E and Dale H. Pretzer. "Comparative Costs of Converting Shelf List Records to Machine Readable Form." *Journal of Library Automation,* v. 1, no. 1, March, 1968, p. 66-74.

(13) *CONVERSION of Retrospective Catalog Records to Machine Readable Form; a Study of the Feasibility of a National Bibliographic Service.* Prepared by the RECON Working Task Force, Henriette D. Avram, Chairman. Washington, D.C., Library of Congress, 1969.

(14) COWARD, R. E., "MARC International." *Journal of Library Automation*, v. 2, no. 4, December 1969, p. 181-186.

(15) COWARD, R.E. "The United Kingdom MARC Record Service" (In *Organization and Handling of Bibliographic Records by Computer*, edited by Nigel S. M Cox and Michael W Grose Newcastle upon Tyne, Oriel Press. 1967, p. 105-115).

(16) CUNNINGHAM, Jay L., William D. Schieber, and Ralph M. Shoffner *A Study of the Organization and Search of Bibliographic Holdings Records in On-Line Computer Systems Phase I Final Report*. Project No. 7-1083, Institute of Library Research, University of California, Berkeley, Calif Washington, D.C., U.S. Dept of Health, Education, and Welfare, Office of Education, Bureau of Research, March 1969. PB 185655

(17) CURRAN, Ann T and Henriette D. Avram *The Identification of Data Elements in Bibliographic Records*. Final Report of the Special Project on Data Elements for the Subcommittee on Machine Input Records (SC-2) of the Sectional Committee on Library Work and Documentation (Z-39) of the United States of America Standards Institute May 1967. AD 666447

(18) DE GENNARO, Richard "Automation in the Harvard College Library". *Harvard Library Bulletin*, v 16, no 3, July 1968, p. 217-236.

(19) DE GENNARO, Richard. "A Computer Produced Shelf List." *College and Research Libraries*, v 26, no. 4, July 1965, p. 311-315. 353.

(20) DE GENNARO, Richard "A Strategy for the Conversion of Research Library Catalogs to Machine Readable Form." *College and Research Libraries*, v 28, no 4, July, 1967, p. 253-257

(21) DOLBY, J. L. and V. J Forsyth. "An Analysis of Cost Factors in Maintaining and Updating Card Catalogs." *Journal of Library Automation*, v. 2, no. 4, December 1969, p 218-241.

(22) DOLBY, James L., V. J Forsyth, and H. L Resnikoff. *Computerized Library Catalogs Their Growth, Cost and Utility*. Cambridge, Mass., The M.I.T Press, 1969

(23) HAMMER, Donald P. "Problems in the Conversion of Bibliographical Data - A Keypunching Experiment." *American Documentation*, v 19, no. 1, January 1968, p 12-17

(24) JOHNS HOPKINS UNIVERSITY. Milton S Eisenhower Library. *Final Report (Number 5) on an Operations Research and Systems Engineering Study of a University Library*. Baltimore, December 1968. PB 182834

(25) JOHNS HOPKINS UNIVERSITY. Milton S Eisenhower Library. *Progress Report on an Operations Research and Systems Engineering Study of a University Library* Baltimore, 1965

(26) JOHNSON, Richard D. "A Book Catalog at Stanford." *Journal of Library Automation*, v 1, no. 1, March 1968, p 13-50

(27) JOLLIFFE, John "The Tactics of Converting a Catalogue to Machine-Readable Form." *Journal of Documentation* v 24, no. 3, September 1968, p 149-158

(28) KILGOUR, Frederick G. "Library Catalogue Production on Small Computers." *American Documentation*, v. 17, no. 3, July 1966, p. 124-131.

(29) KRINOS, J. D. *Conversion of Existing Library Catalogues to Computer Files.* Farmington, Conn., Hamilton Standard, Division of United Aircraft Corporation, 12 January, 1968 (SP 01U68) Presented at the fifth National Colloquium on Information Retrieval, Philadelphia, 4 May 1968

(30) LIBRARY OF CONGRESS. Information Systems Office. *MARC Manual. V. 2, Data Preparation Manual: MARC Editors.* December, 1968. Chicago, American Library Association, 1969.

(31) LIBRARY OF CONGRESS. Information Systems Office. *Subscriber's Guide to the MARC Distribution Service. Specifications for Magnetic Tapes Containing Monographic Catalog Records in the MARC II Format* Washington D. C., 1968

(32) MASSACHUSETTS INSTITUTE OF TECHNOLOGY. Project INTREX. *Semiannual Activity Report.* 15 March 1968 to 15 September 1968. Cambridge, Mass 1968

(33) NIEWALDA, Paul, and Gertrud Preuss. "Die Elektronik im Dienste der Katalogisierung der Universitätsbibliothek Regensburg; ein Erfahrungsbericht" *Zeitschrift für Bibliothekswesen und Bibliographie,* v. 16, no. 2, 1969, p. 86-118.

(34) *OPTICAL Character Recognition Research and Demonstration Project* Los Angeles County Public Library, 1968. ED 019974

(35) PARKER, Patricia E. "The Preparation of MARC Bibliographic Data for Machine Input." *Library Resources and Technical Services,* v 12, no. 3, Summer 1968, p. 311-319.

(36) PAYNE, Charles. "The University of Chicago's Book Processing System." *Stanford Conference on Collaborative Library Systems Development, Proceedings of a Conference Held at Stanford University Libraries, October 4-5, 1968.* Stanford, Ca., 1969. p. 119-127, 134-143. ED 031281

(37) PERREAULT, Jean M. "The Computer Book Catalog at Florida Atlantic University" *College & Research Libraries,* v. 25, no. 3, May 1964. p. 185-197.

(38) RICHMOND, Phyllis A "A Short-Title Catalog Made with IBM Tabulating Equipment." *Library Resources and Technical Services,* v. 7, no. 1, Winter 1963. p. 81-90

(39) SIMMONS, P. A "An Analysis of Bibliographic Data Conversion Costs" *Library Resources & Technical Services,* v. 12, no. 3, Summer 1968, p. 296-311

(40) SOBOTTKE, Ingeborg "Rationalisierung der alphabetischen Katalogistierung." (In *Elektronische Datenverarbeitung in der Universitätsbibliothek Bochum,* hrsg von Günther Pflug und Bernhard Adams. Bochum, 1968. p.24-32)

(41) SOUTHERN ILLINOIS UNIVERSITY. Office of Systems and Procedures *An Automated Circulation Control System for the Delyte W. Morris Library, the System and Its Progress in Brief.* Carbondale, Ill., Southern Illinois University, 1963

(42) STOLTZENBURG, Joachim, and Günther Rabe "Das Konstanzer Schema für Erfassung bibliographischer Daten." *Zeitschrift für Bibliothekswesen und Bibliographie,* v 16, no. 2, 1969, p. 119-153.